KELLOGG, HANSEN, TODD, FIGEL & FREDERICK, P.L.L.C.
SUMNER SQUARE
1615 M STREET, N.W.
SUITE 400
WASHINGTON, D.C. 20036-3215

(202) 326-7900
FACSIMILE:
(202) 326-7999

November 8, 2024

Via ECF

Hon. William H. Alsup
United States District Court for the Northern District of California
Courtroom 12, 19th Floor
450 Golden Gate Ave
San Francisco, CA 94102

Re: *X Corp. v. Bright Data Ltd.*, No. 23-cv-03698 (N.D. Cal.)

Dear Judge Alsup:

We write in response to Bright Data's October 30, 2024 letter previewing its proposed motion to dismiss the three new statutory claims in X's Second Amended Complaint ("SAC"). Dkt. 141. Bright Data's latest arguments do not warrant filing another motion to dismiss, especially after it chose not to challenge the merits of these claims when opposing leave to amend. In any event, Bright Data's arguments raise factual disputes that cannot be resolved on the pleadings. *E.g.*, *Tippitt v. Life Ins. Co. of N. Am.*, 2017 WL 3189464, at *1 (N.D. Cal. May 30, 2017) ("A motion to dismiss is not an appropriate mechanism for resolving factual disputes."). Taken as true, the SAC's allegations state a claim under these statutes for the reasons X explained when seeking leave to amend. Dkt. 118 at 19-23. The parties should thus focus on finally moving discovery forward, not on Bright Data's meritless motion to dismiss.

1. X Plausibly Alleges That Bright Data and Its Customers Scrape Password-Protected Data and Circumvent X's Technological Access Restrictions

At the threshold, Bright Data repeatedly insists it has not violated the three statutes at issue – the Computer Fraud and Abuse Act ("CFAA"), California's Comprehensive Computer Data and Access Fraud Act ("CDAFA"), and the Digital Millennium Copyright Act ("DMCA") – because it accesses only "public" data. That argument fails for several reasons.

First, X plausibly alleges that Bright Data has scraped data behind a log-in screen. SAC ¶¶ 3, 87, 105, 120, 205, 215. Indeed, the SAC specifically "demonstrates that most data scraping occurs through logged-in accounts . . . , *not* through logged-out scraping of public data[,]" particularly for web requests to view an individual user's profile, followers and followings, posts and replies, and post timeline. *Id.* ¶ 87. The Court must accept these allegations as true. And

Bright Data does not dispute that, if it engaged in logged-in scraping, it violated each statute. The Court can dispatch all of Bright Data's arguments on that ground alone.

Second, even for "specific post[s]," the "number of likes, replies and reposts," the "profile of the person who posted the content," "a curated list of other posts from that individual," and other "linked posts," Dkt. 141 at 1, the SAC makes clear that a logged-out user's ability to view this content "is both rate limited and qualitatively restricted," SAC ¶ 67. Accordingly, "there is no way Bright Data could have obtained [the high volume of data it sells] without circumventing the technological measures X put in place to prevent scraping of data available to users that are not logged into X." *Id.* ¶ 120. As described below, Bright Data's circumvention of X's technological access restrictions violates the CFAA, CDAFA, and DMCA.

Third, even if Bright Data scrapes only publicly accessible data, that would not defeat the claims. Indeed, DMCA liability does not turn on whether the scraped data is public or private. The question is whether Bright Data circumvented technological measures controlling access to X's platform, *see* 17 U.S.C. § 1201(a)(1)(A), or sold products primarily designed, produced, or marketed for this purpose, *id.* § 1201(a)(2). The SAC alleges it did. SAC ¶¶ 189-199. Similarly, the CDAFA asks whether a defendant *uses* data without permission – not whether the *access* was authorized – and X alleges that Bright Data lacked permission to use the data it scraped. *Id.* ¶¶ 30-35, 58-79. Finally, even if discovery were to reveal that Bright Data itself does not scrape data only available to logged-in users – which cannot be resolved on a motion to dismiss – X alleges that Bright Data advertises and sells tools that allow its customers to do so. *Id.* ¶¶ 65 (conversation threads accessible only to logged-in users); 124(a) (Bright Data's Twitter Scraper scrapes data from conversation threads). And Bright Data does not dispute that it may be vicariously liable under the CFAA for its customers' violations. *See Facebook, Inc. v. Power Ventures, Inc.*, 844 F.3d 1058, 1066-67 (9th Cir. 2016); SAC ¶¶ 105, 120-134, 206.

2. The SAC Plausibly Alleges Violations of the CFAA and CDAFA

The CFAA prohibits, among other things, "intentionally access[ing] a computer without authorization or exceed[ing] authorized access, and thereby obtain[ing] . . . information from any protected computer." 18 U.S.C. § 1030(a)(2)(C). The CDAFA imposes liability on any person who "[k]nowingly accesses and without permission takes, copies, or makes use of any data from a computer, computer system, or computer network." Cal. Penal Code § 502(c)(2).

Bright Data does not address X's CFAA and CDAFA claims separately. Rather, it argues (at 3) that X's CDAFA claim "fails for the same reasons" as its CFAA claim. Although the elements of the CFAA and CDAFA align closely, Bright Data glosses over an important distinction. The CDAFA "does not require *unauthorized* access. It merely requires *knowing* access." *United States v. Christensen*, 828 F.3d 763, 789 (9th Cir. 2015) (explaining that "access" for purposes of the CDAFA "includes logging into a database with a valid password and subsequently taking, copying, or using the information in the database improperly"). Accordingly, CDAFA liability focuses on whether a defendant's *use* of data is "without permission," not whether the defendant's access to the data lacked authorization.

Bright Data elides that distinction in challenging X's CFAA and CDAFA claims on only one ground: that it did not access any data "without authorization" because it accessed only public data. Again, that argument fails because Bright Data's factual assertion that the scraped data was "public" ignores X's well-pleaded allegations. *See supra* p. 1.

In any event, the Ninth Circuit has held that a scraper accesses a computer system "without authorization" when it "circumvents a computer's generally applicable rules regarding access permissions, such as username and password requirements, to gain access." *hiQ Labs, Inc. v. LinkedIn Corp*., 31 F.4th 1180, 1201 (9th Cir. 2022). X plausibly alleges that Bright Data has done just that by (among other things) bypassing IP rate limits that redirect users to the log-in page, creating fake X accounts to circumvent X's CAPTCHA prompts, and circumventing robots.txt. SAC ¶¶ 1, 5, 25-28, 64-79, 105-34.[1] Such allegations also show that Bright Data knowingly accessed data on X's platform and used that data "without permission" under the CDAFA. *See NovelPoster v. Javitch Canfield Grp*., 140 F. Supp. 3d 938, 950 (N.D. Cal. 2014) (parties act "without permission" when they circumvent technical or code-based barriers limiting user access); *Craigslist, Inc. v. Naturemarket, Inc*., 694 F. Supp. 2d 1039, 1057 (N.D. Cal. 2010) (website operator stated claim against software developer where developer knowingly accessed its computer system by bypassing similar access controls).[2]

3. X Plausibly Alleges a Violation of the DMCA

The DMCA bars "circumventing a technological measure that effectively controls access to" a copyrighted work. 17 U.S.C. § 1201(a)(2). A related provision bars "(1) traffic[king] in (2) a technology or part thereof (3) that is primarily designed, produced, or marketed for, or has limited commercially significant use other than (4) circumventing a technological measure (5) that effectively controls access (6) to a copyrighted work." *MDY Indus., LLC v. Blizzard Ent., Inc*., 629 F.3d 928, 953 (9th Cir. 2010). There is no dispute that X's websites and mobile app are protected under the Copyright Act. Bright Data argues only (at 4-5) that X's anti-scraping safeguards are not "technological measure[s] that effectively control[] access" to its platform, and that it did not circumvent these measures. Both assertions lack merit.

Bright Data's threshold argument is (again) that X's technological measures do not "control access to X's copyright content" because the content is publicly available. Dkt. 141 at 4. But whether the data Bright Data and its customers scraped is "public" makes no difference to

[1] Bright Data cites (at 3) *Meta Platforms, Inc. v. BrandTotal Ltd*., 605 F. Supp. 3d 1218 (N.D. Cal. 2022). But in that case, the court granted Meta summary judgment because the evidence showed the defendant had "used 'fake' user credentials to access restricted pages." *Id.* at 1267. That is one of the things X alleges Bright Data did here. *E.g.*, SAC ¶ 79.

[2] *Oracle USA, Inc. v. Rimini Street, Inc*., 879 F.3d 948 (9th Cir. 2018), does not hold otherwise. In *Oracle*, the user "was authorized in the first instance to take and use the information that it downloaded." *Id.* at 962. The Ninth Circuit held that, where taking the data is permitted and the only issue is the method of taking, the CDAFA does not apply. But here, Bright Data did *not* have permission to access the data it scraped and instead circumvented numerous "technical [and] code-based barriers" to do so, all in violation of X's terms of service. *NovelPoster*, 140 F. Supp. 3d at 950.

the DMCA claim. The only question is whether Bright Data and its customers circumvented X's technological measures when scraping such data. The SAC pleads they did so in many ways. Indeed, all the most valuable data on X's platform – even the data (like user posts) that is theoretically "public" – is locked behind technological measures. SAC ¶¶ 3, 25-28, 64-79.

Bright Data Circumvented X's Technological Measures. To start, Bright Data does not dispute that it circumvents X's robots.txt files. SAC ¶¶ 75-78, 191. Rather, it asserts (at 4-5) that it did not circumvent X's CAPTCHAs, login requirements, or rate limits. That factual assertion ignores allegations in the SAC and cannot be resolved on a motion to dismiss.

X employs CAPTCHAs to ensure that a human (rather than a bot) is creating an account. SAC ¶¶ 27-28, 70. X pleads that Bright Data sells tools that automatically crack CAPTCHA, *id.* ¶¶ 123, 130-31, 193, and that it circumvents X's CAPTCHAs to automatically open legions of fake accounts to access otherwise-restricted data on X's platform, *id.* ¶¶ 79, 93. These allegations are sufficient to plead a DMCA claim. *E.g.*, *Ticketmaster L.L.C. v. Prestige Ent. W., Inc.*, 315 F. Supp. 3d 1147, 1166-67 (C.D. Cal. 2018) (sustaining DMCA claim against data scraper for using bots to circumvent "security measures, including CAPTCHA" that "control access to Ticketmaster's copyrighted webpages"); *Craigslist, Inc. v. Kerbel*, 2012 WL 3166798, at *10 (N.D. Cal. Aug. 2, 2012) (denying dismissal of DMCA claim where defendant marketed auto-posting services that enabled users to circumvent Craigslist's CAPTCHA).

For log-in requirements, Bright Data argues (at 4) that it cannot be liable under the DMCA because it never entered "fraudulent passwords." X alleges, however, that Bright Data and its customers "can only access the content they want through logged-in accounts, so [they] use fake, automatically created accounts to obtain it." SAC ¶ 3; *see also id.* ¶¶ 79, 132. Courts have found that the application of fraudulent credentials is circumvention that violates the DMCA. *See Synopsys, Inc. v. InnoGrit, Corp.*, 2019 WL 4848387, at *8-9 (N.D. Cal. Oct. 1, 2019) (use of real, but unauthorized, license key to activate software was circumvention) (collecting cases); *Microsoft Corp. v. EEE Bus. Inc.*, 555 F. Supp. 2d 1051, 1059 (N.D. Cal. 2008) (similar); *see also RealNetworks, Inc. v. Streambox, Inc.*, 2000 WL 127311, at *4, 7 (W.D. Wash. Jan. 18, 2000) (finding circumvention where "the Streambox VCR is able to convince the RealServer into thinking that the VCR is, in fact, a RealPlayer").[3]

Similarly, X uses IP-based rate limits to control logged-out scraping. SAC ¶¶ 67-68, 71-71. Thus, data scraping on X's platform can occur only "through the coordinated use of proxy

[3] *iSpot.tv, Inc. v. Teyfukova*, 2023 WL 1967958 (C.D. Cal. Jan. 25, 2023), which Bright Data cites (at 4), similarly recognized that "if iSpot alleged that Teyfukova had improperly obtained the username and password – either through deception or through some more technological means – a question might be raised as to whether this constitutes 'avoiding' or 'bypassing' the system," *id.* at *12. The other cases cited by Bright Data are inapposite. *See Aeropost Int'l Servs., Inc. v. Aerocasillas, S.A.*, 2011 WL 13174672, at *6 (S.D. Fla. Mar. 31, 2011) (plaintiff did not allege any technological impediment to creating additional logins and passwords); *Schork Grp., Inc. v. Choice! Energy Servs., Retail, LP*, 2022 WL 2905231, at *10 (E.D. Pa. July 21, 2022) (plaintiff failed to detail the measures used to circumvent access and provided defendant "with a legitimate password to access the" content at issue).

networks with massive amounts of rotating IP addresses (which Bright Data provides)." *Id.* ¶ 94. "These proxy services imitate requests from legitimate users to conceal the true requestor's IP address and location." *Id.* ¶ 131. Bright Data says at (5) it did not circumvent X's rate limits because it merely "switch[ed] IP addresses" to "restart[] the rate limiter." But spoofing an IP address to fool (*i.e.*, avoid, deactivate, or impair) an IP rate limiter is textbook circumvention, and courts consistently hold that similar efforts to "spoof" a device's characteristics to "trick" and therefore "get around" a technological measure violates the DMCA. *E.g.*, *Ticketmaster L.L.C. v. Prestige Ent., Inc.*, 306 F. Supp. 3d 1164, 1174 (C.D. Cal. 2018) ("Defendants' use of colocation facilities and other methods, such as deleting tracking tools like 'cookies,' are also actionable . . . if used to circumvent Ticketmaster's technological measures."); *Synopsys, Inc. v. InnoGrit, Corp.*, 2019 WL 2617091, at *3 (N.D. Cal. June 26, 2019) (holding that changing the MAC addresses on computers to run unauthorized software is circumvention).

X's Technological Measures Effectively Control Access to its Platform. Finally, Bright Data argues (at 5) that neither X's rate limits nor its robots.txt files are "effective controls." This too is a factual question the Court cannot resolve on motion to dismiss.

Bright Data says (at 5) that rate limits are not access controls because they "do not prevent access to the website; just the frequency with which such information can be accessed." But when a user reaches the limit, X prevents that user from further *accessing* the site – that is the whole point of the limit. SAC ¶ 71. To access X's service, users must provide the information relied on by X's rate limiter, including the user's IP address, identifying information for the web browser, and the actions the user wishes to perform on X. *Id.* ¶ 74. The rate limiter thus requires application of a process (the rate limit) and information (about the user) to access X's website. When a user fails this test, X redirects the user to the login page and bars them from further accessing the site. That is all Section 1201(a)(3)(B) requires.

Similarly, X's robots.txt files prevent bots – except for Google's web crawler – from automated access to its website. This measure is not "voluntary," as Bright Data claims; rather, "X Corp.'s robots.txt instructions explicitly forbid any scraping of its website" by Bright Data and its customers. SAC ¶ 75. Courts have found robots.txt files qualify as "effective controls" under similar circumstances. *E.g.*, *DHI Grp., Inc. v. Kent*, 2017 WL 8794877, at *6 (S.D. Tex. Apr. 21, 2017), *report and recommendation adopted*, 2017 WL 4837730 (S.D. Tex. Oct. 26, 2017) (allegations that plaintiff used a robots.txt file to prevent automated technologies from accessing the website sufficient to state a DMCA claim). Indeed, even the one case Bright Data cites (at 1 n.1), *Healthcare Advocates, Inc. v. Harding, Earley, Follmer & Frailey*, 497 F. Supp. 2d 627, 643 (E.D. Pa. 2007), found that a similar robots.txt file was a technological measure effectively controlling access to its website. The same is true here.

Respectfully submitted,

/s/ Joshua D. Branson

Joshua D. Branson